FIGURE 1: Annotation of GAS 40
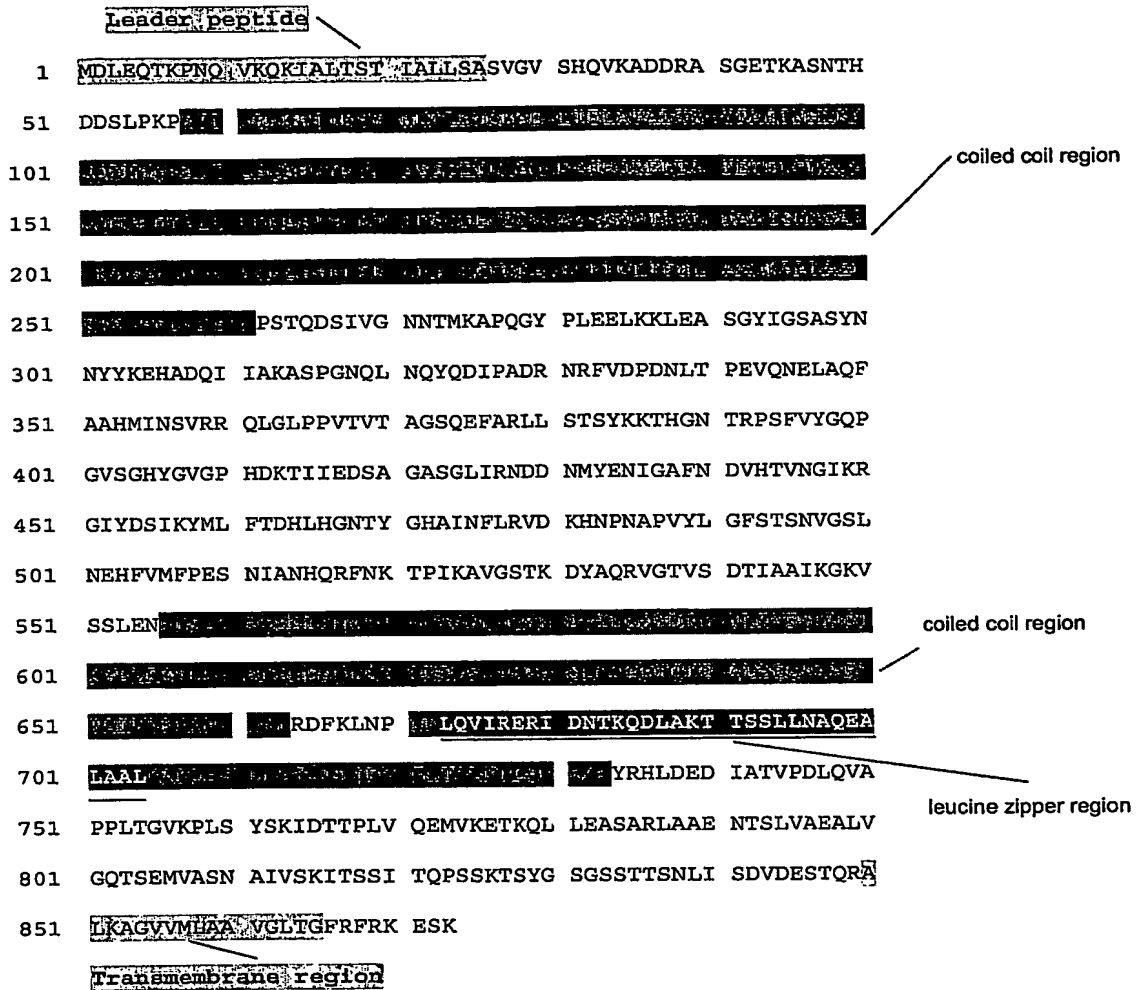
Leader peptide

```
  1  MDLEQTKPNQ VKQKIALTST IAILSASVGV SHQVKADDRA SGETKASNTH
 51  DDSLPKP▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓      coiled coil region
101  ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓
151  ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓
201  ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓
251  ▓▓▓▓▓▓▓▓▓PSTQDSIVG NNTMKAPQGY PLEELKKLEA SGYIGSASYN
301  NYYKEHADQI IAKASPGNQL NQYQDIPADR NRFVDPDNLT PEVQNELAQF
351  AAHMINSVRR QLGLPPVTVT AGSQEFARLL STSYKKTHGN TRPSFVYGQP
401  GVSGHYGVGP HDKTIIEDSA GASGLIRNDD NMYENIGAFN DVHTVNGIKR
451  GIYDSIKYML FTDHLHGNTY GHAINFLRVD KHNPNAPVYL GFSTSNVGSL
501  NEHFVMFPES NIANHQRFNK TPIKAVGSTK DYAQRVGTVS DTIAAIKGKV
551  SSLEN▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓      coiled coil region
601  ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓
651  ▓▓▓▓▓▓▓▓▓ ▓▓▓RDFKLNP ▓▓LQVIRERI DNTKQDLAKT TSSLLNAQEA
701  LAAL▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓▓▓▓▓▓▓▓ ▓▓YRHLDED IATVPDLQVA
751  PPLTGVKPLS YSKIDTTPLV QEMVKETKQL LEASARLAAE NTSLVAEALV
801  GQTSEMVASN AIVSKITSSI TQPSSKTSYG SGSSTTSNLI SDVDESTQRA
851  LKAGVVMHAA VGLTGFRFRK ESK
```

leucine zipper region

Transmembrane region

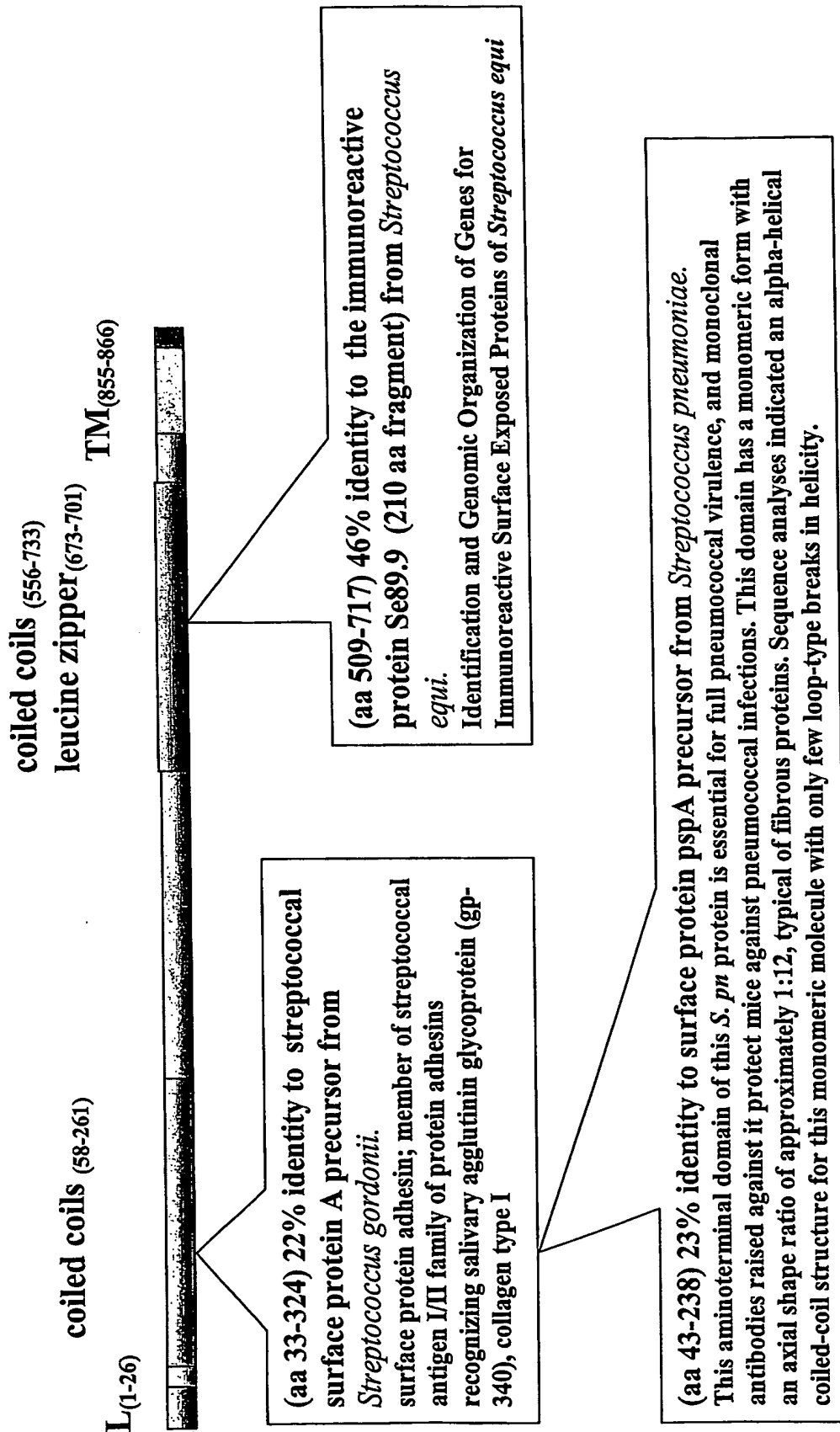# FIGURE 2 : Schematic of GAS40: putative surface exclusion protein prgA (873aa)

coiled coils (556-733)
leucine zipper(673-701) TM(855-866)

coiled coils (58-261)

L(1-26)

(aa 33-324) 22% identity to streptococcal surface protein A precursor from *Streptococcus gordonii.* surface protein adhesin; member of streptococcal antigen I/II family of protein adhesins recognizing salivary agglutinin glycoprotein (gp-340), collagen type I

(aa 509-717) 46% identity to the immunoreactive protein Se89.9 (210 aa fragment) from *Streptococcus equi.*
Identification and Genomic Organization of Genes for Immunoreactive Surface Exposed Proteins of *Streptococcus equi*

(aa 43-238) 23% identity to surface protein pspA precursor from *Streptococcus pneumoniae.*
This aminoterminal domain of this *S. pn* protein is essential for full pneumococcal virulence, and monoclonal antibodies raised against it protect mice against pneumococcal infections. This domain has a monomeric form with an axial shape ratio of approximately 1:12, typical of fibrous proteins. Sequence analyses indicated an alpha-helical coiled-coil structure for this monomeric molecule with only few loop-type breaks in helicity.

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(a) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with SpA precursor of *Streptococcus gordonii*

```
>gi|25990270|gb|AAC44101.3| streptococcal surface protein A precursor
[Streptococcus gordonii]
         Length = 1575
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus pyogenes]
         Length = 873

 Score = 63.2 bits (152), Expect = 5e-11
 Identities = 65/293 (22%), Positives = 124/293 (42%), Gaps = 13/293
(4%)

Query: 112 QDQTSDKGTATTAAENAQKQAEIKSDYAKQA----EEIKKTTEAYKKEVEAHQAETDKIN
167
           Q +  D+ +  T A N    + K +  ++A    + ++KT    K E+         K
Sbjct: 33  QVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVEKTLSQQKAELTELATALTKTT
92

Query: 168 AENKAAEDKYQEDLKAHQAEVEKINTANATAKAEYEAKLAQYQKDLAAVQKANEDSQLDY
227
           AE    +++   + KA + E       A+++   A+ A++Q++L A +     ++Q D
Sbjct: 93  AEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATETELHNAQADQ
152

Query: 228 QNKLSAYQAELARVQKANAEAKEAYE--KAVKENTAKNAALQAENEAIKQRNETAKANYD
285
           +K +A   + A +        A++  E K  ++N AK  A+ +   +AI + +TA  N
Sbjct: 153 HSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTK
212

Query: 286 AAMKQYEADLAAIKKAKEDNDADYQAKLAAYQAELARVQKANADAKAAYEKAVEENTAKN
345
           A  + E  A ++  K          +LAA +A LA +   +   K++      +     N
Sbjct: 213 ALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSIVGNN
272

Query: 346 TAIQAEN---EAIKQRNAA---AKATYEAALKQYEADLAAAKKANEDSDADYQ 392
           T   +    E +K+ A+       A+Y    K++ AD  AK +  +   YQ
Sbjct: 273 TMKAPQGYPLEELKKLEASGYIGSASYNNYYKEH-ADQIIAKASPGNQLNQYQ 324
```

**FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other
Streptococcus bacteria**

**3(b) BLAST alignment of amino acid sequence of GAS 40 including the first
coiled-coil region with SpB precursor of *Streptococcus gordonii***

```
>gi|25055226|gb|AAC44102.3|    streptococcal surface protein B precursor
[Streptococcus gordonii]
          Length = 1499
>ref|NP 268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
          Length = 873


 Score = 54.3 bits (129), Expect = 2e-08
 Identities = 53/226 (23%), Positives = 98/226 (43%), Gaps = 13/226
(5%)


Query: 111 QDQTSDKGTATTAAENAQKQAEIKSDYAKQA----EEIKKTTEAYKKEVEAHQAETDKIN
166
           Q + D+ + T A N  + K +  ++A   + ++KT    K E+         K
Sbjct: 33  QVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAVEKTLSQQKAELTELATALTKTT
92


Query: 167 AENKAAEDKYQEDLKAHQAEVEKINTANATAKAEYEAKLAQYQKDLAAVQKANEDSQLDY
226
           AE    +++   + KA + E     A+++    A+ A++Q++L A +    ++Q D
Sbjct: 93  AEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATETELHNAQADQ
152


Query: 227 QNKLSAYQAELARV--QXXXXXXXXXXXXXXXXXXNTAKNAALQAENEAIKQRNETAKANYD
284
           +K +A   + A +  +              N AK A+ +  +AI +   +TA  N
Sbjct: 153 HSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTK
212


Query: 285 AAMKQYE---ADL----AAIKKAKEDNDADYQAKLAAYQAELARVQ 323
           A   + E   ADL    A +KK   + A  +A LA  +AEL+R++
Sbjct: 213 ALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLK 258
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(c) BLAST alignment of amino acid sequence of GAS 40 including the first coiled-coil region with Surface Protein PspA precursor of _Streptococcus pneumoniae_

```
>gi| 282335 |pir||A41971 surface protein pspA precursor - Streptococcus
pneumoniae
>ref|NP 268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
          Length = 873

 Score = 48.1 bits (113), Expect = 6e-07
 Identities = 46/200 (23%), Positives = 89/200 (44%), Gaps = 23/200
(11%)

Query: 139 KTKFNTVRAMVVPEPEQLAETK-------KKSEEAKQKAPELTKKLEEAKAKLEE-AEKK
190
           +TK +       +P+PE + E K        K   + K + EL   L +  A++     E++
Sbjct: 43  ETKASNTHDDSLPKPETIQEAKATIDAVEKTLSQQKAELTELATALTKTTAEINHLKEQQ
102

Query: 191 ATEAKQKVDAEEVAPQAKIAELENQVHRLEQELKEIDESESEDYAKEGFRAPLQSKLDAK
250
            E K    A+E+       + E + +  +   +E+   +E+E +  +  +    ++ L   +
Sbjct: 103 DNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTATETELHNAQADQHSKETALSEQ
162

Query: 251 KAKLS----KLEELSDKIDELDAEIAKLEDQL-------KAAEENNNVEDYFKEGLEKTI
299
           KA +S     + ++L +++    + IAKL   +        KAA+ N+        LEK
Sbjct: 163 KASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDNTKALSSELEKA-
221

Query: 300 AAKKAELEKTEADLKKAVNE 319
               KA+LE  +A +KK + E
Sbjct: 222 ---KADLENQKAKVKKQLTE 238
```

FIGURE 3: BLAST results of Coiled-Coil regions of GAS 40 with other Streptococcus bacteria

3(d) BLAST alignment of amino acid sequence of GAS 40 including the second coiled-coil region with SpB precursor of *Streptococcus gordonii*

```
>gi|23380384|gb|AAN18299.1|    immunoreactive protein Se89.9 (fragment)
[Streptococcus equi]
        Length = 210
>ref|NP_268623.1| putative surface exclusion protein [Streptococcus
pyogenes]
        Length = 873

 Score =  173 bits (438), Expect = 4e-45
 Identities = 98/209 (46%), Positives = 144/209 (68%)


Query: 1    ESDIVDATRFSTTEIPKSGQVIDRSASIQALTNDIASIKGKIASLESRLADPSSEAEVTA
60
            ES+I +  RF+ T I   G   D +  +  +++ IA+IKGK++SLE+RL+    EA++ A
Sbjct: 509  ESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKVSSLENRLSAIHQEADIMA
568


Query: 61   AQAKISQLQHQLEAAQAKSHKLDQQVEQLANTKDSLRTQLLAAKEEQAQLKANLDKALAL
120
            AQAK+SQLQ +L +   +S  L+ QV QL +TK SLRT+LLAAK +QAQL+A  D++LA
Sbjct: 569  AQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAK
628


Query: 121  LASSKATLHKLEAAMEEAKARVAGLASQKAQLEDLLAFEKNPNRIELAQEKVAAAKKALA
180
            LAS KA LH+ EA  E+A ARV  L ++KA L+ L  F+ NPNR+++ +E++    K+ LA
Sbjct: 629  LASLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLA
688


Query: 181  DTEDKLLAAQASLSDLQAQRARLQLSIAT 209
                T    LL AQ +L+ LQA+++ L+ +IAT
Sbjct: 689  KTTSSLLNAQEALAALQAKQSSLEATIAT 717
```

Figure 4: Secondary Structure Prediction of GAS 40

Figure 4(a) Secondary Structure prediction alignment with GAS 40 amino acid sequence

```
          10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MDLEQTKPNQVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAV
CCCCCCCChhhHHhhhhhHHHHhhhccceeEEEecCCcCCCCCcCCCCCCCCCCCCcHHHHHHHHHHHH
EKTLSQQKAELTELATALTKTTAEINHLKEQQDNEQKALTSAQEIYTNTLASSEETLLAQGAEHQRELTA
HHHHHHHHHHHHHHHHHHHHhhHHHHHHHHHhhhHHHHHHHHHHHHHHHhccchHHHHHHHHHHHHHHHHH
TETELHNAQADQHSKETALSEQKASISAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDN
HHHHHHHHHHHcccchhHHHHHhhhhccehhhhHHHHHHHHHHhhHHHHHHHHHHhhhcCcHHHHHHHHHHhhc
TKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEAELSRLKSSAPSTQDSIVGNNTMKAPQGY
cHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhcCCCCCCCCeEcCCCCCCCCCC
PLEELKKLEASGYIGSASYNNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPDNLTPEVQNELAQF
CHHHHHHHhcCccceeccchHHHHHHHHHHHHHHHhCCchhhhhhccCccccCCCCCCCCChHHHHHHHHH
AAHMINSVRRQLGLPPVTVTAGSQEFARLLSTSYKKTHGNTRPSFVYGQPGVSGHYGVGPHDKTIIEDSA
HHHHHHHHHHHHcCCCCceecCCCHHHHHHHHHhhcccccCCCCCceEEEcCCCceeecceCcCCCeEEEEcC
GASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRVDKHNPNAPVYL
CCCceecCCcHHHhhhcccccccccCcccccHHHHHHHHhheeecccCccchhHHHheeeeecCCCCCCcEEE
GFSTSNVGSLNEHFVMFPESNIANHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGKVSSLENRLSAI
EEEecCccCcccceecccccccchHHhhhCCCCCcccCCcHHHHHHHchhHHHHHHHHhcCcccHHHHHHHH
HQEADIMAAQAKVSQLQGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLA
HHHHHHHHHHHHHHHHHHHhHHHHHhhccCCchhHHhhhhhcCcCHHHHHHHHHHHHHHHHHHHHhhHHHHHHHH
SLKAALHQTEALAEQAAARVTALVAKKAHLQYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEA
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcCCcHHHHHHHHHHhHHHHHHHHHHHHHHHHHH
LAALQAKQSSLEATIATTEHQLTLLKTLANEKEYRHLDEDIATVPDLQVAPPLTGVKPLSYSKIDTTPLV
HHHHHHhhcCceeecccchHHHHHHHHHHHHHhhhhhhHHHhhccCCCccCCCCCCcCCCceeccCCCHHH
QEMVKETKQLLEASARLAAENTSLVAEALVGQTSEMVASNAIVSKITSSITQPSSKTSYGSGSSTTSNLI
HHHHHHHHHHHHHHHHHHHhHHHHHHHHHhcchhHHHHhhchhhhcceEEecCCCcccccccCccccCcce
SDVDESTQRALKAGVVMLAAVGLTGFRFRKESK
cCCchHHHHHHHhcceeeEeecccccceeeccCC
```

```
Sequence length :    873
PHD :
    Alpha helix      (Hh) :   525 is   60.14%
    3₁₀ helix        (Gg) :     0 is    0.00%
    Pi helix         (Ii) :     0 is    0.00%
    Beta bridge      (Bb) :     0 is    0.00%
    Extended strand  (Ee) :    63 is    7.22%
    Beta turn        (Tt) :     0 is    0.00%
    Bend region      (Ss) :     0 is    0.00%
    Random coil      (Cc) :   285 is   32.65%
    Ambigous states  (?)  :     0 is    0.00%
```
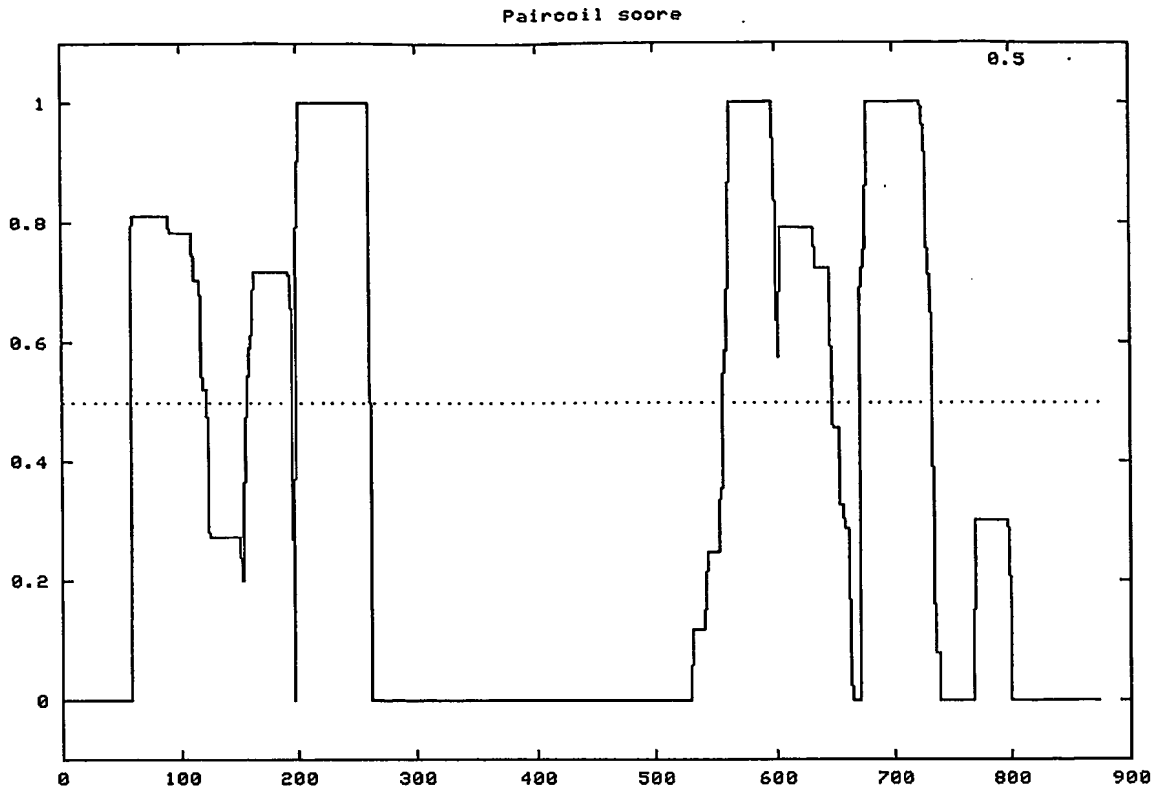
Figure 4(b): Secondary Structure prediction based on PairCoil Score

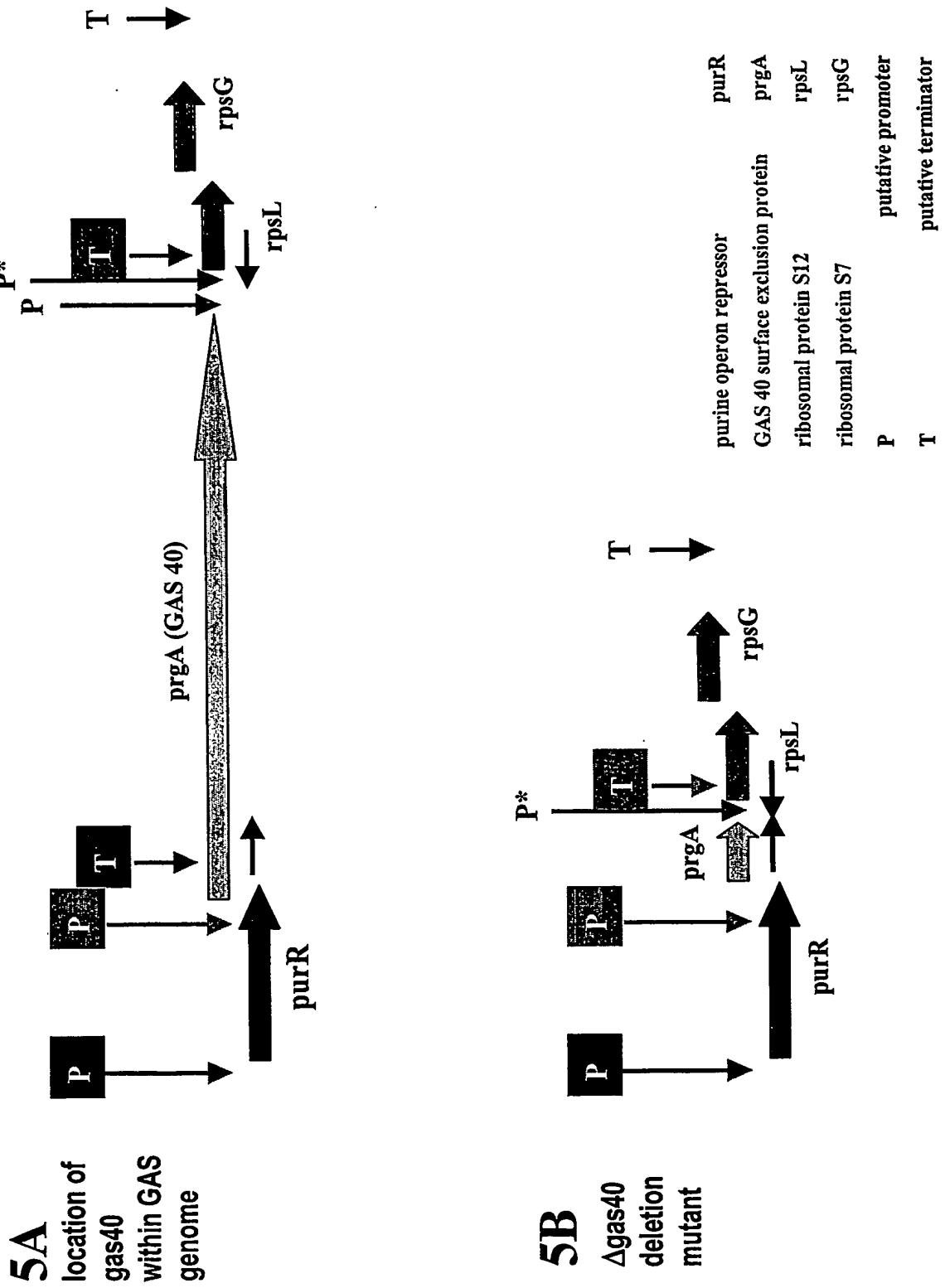Coiled coils found:

```
positions    58- 121    initial register 'f'    probability  0.811
positions   156- 196    initial register 'e'    probability  0.720
positions   198- 246    initial register 'e'    probability  1.000
positions   247- 261    initial register 'a'    probability  1.000

positions   556- 646    initial register 'e'    probability  1.000
positions   671- 733    initial register 'f'    probability  1.000
```
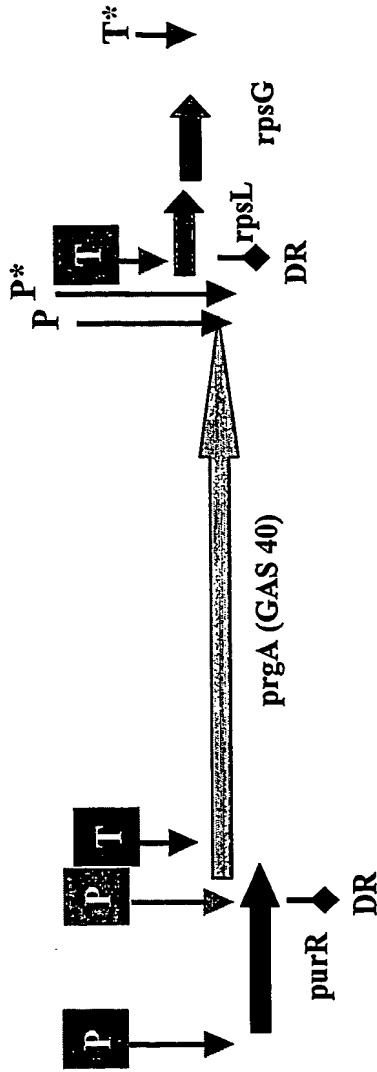
Figure 4(c): Secondary Structure prediction of Leucine Zipper within coiled coil.

```
         673                                    701
QYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEALAALQAKQSSLEATIATTEH
         CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
         L------I------L------L------L
         OOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO
```
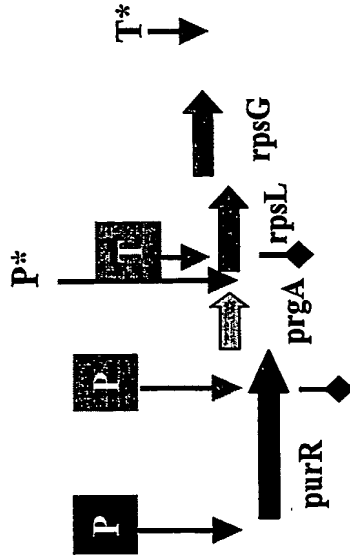
Figure 5

5A
location of
gas40
within GAS
genome

5B
Δgas40
deletion
mutant

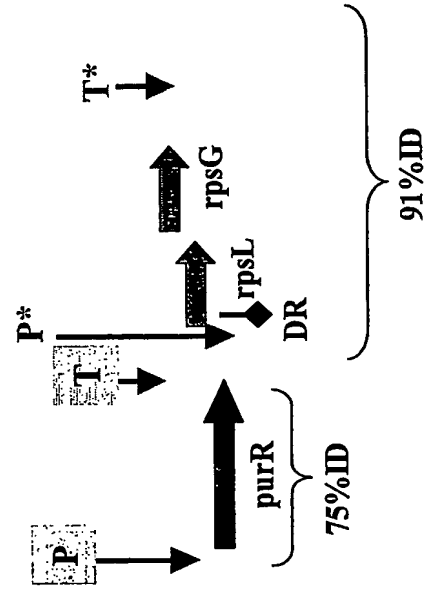|	
purR	purine operon repressor
prgA	GAS 40 surface exclusion protein
rpsL	ribosomal protein S12
rpsG	ribosomal protein S7
P	putative promoter
T	putative terminator

5C

direct repeats surrounding GAS 40 within wild type genome

5D

location of direct repeats within Δgas40

5E

corresponding genomic region within GBS

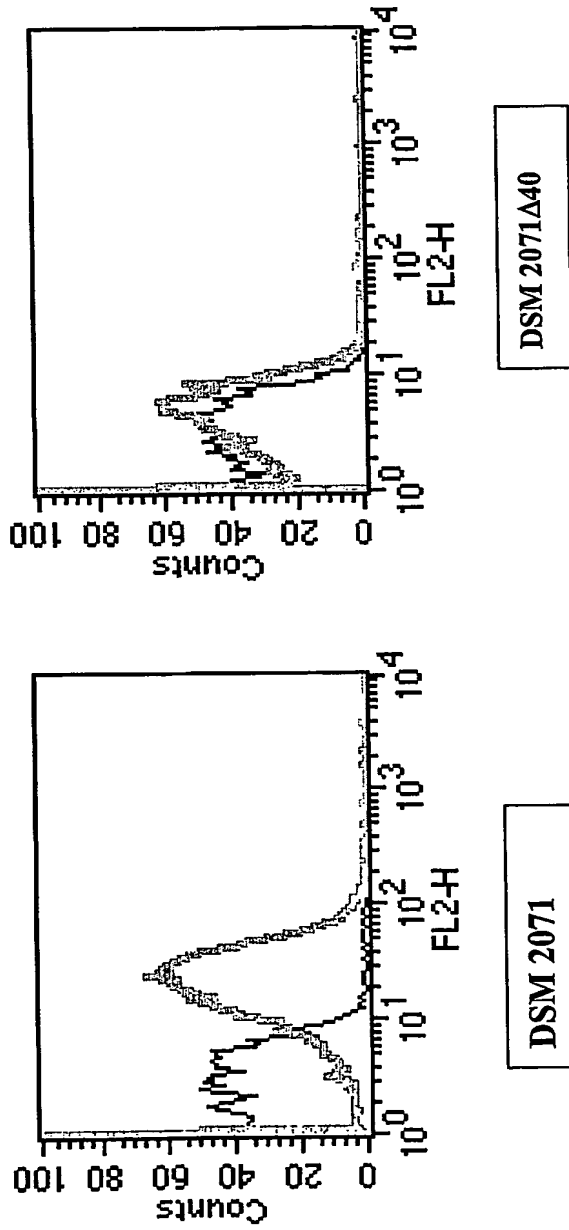**FIGURE 6: FACS Comparison of GAS 40 in wild type GAS and GAS 40 deletion mutant**
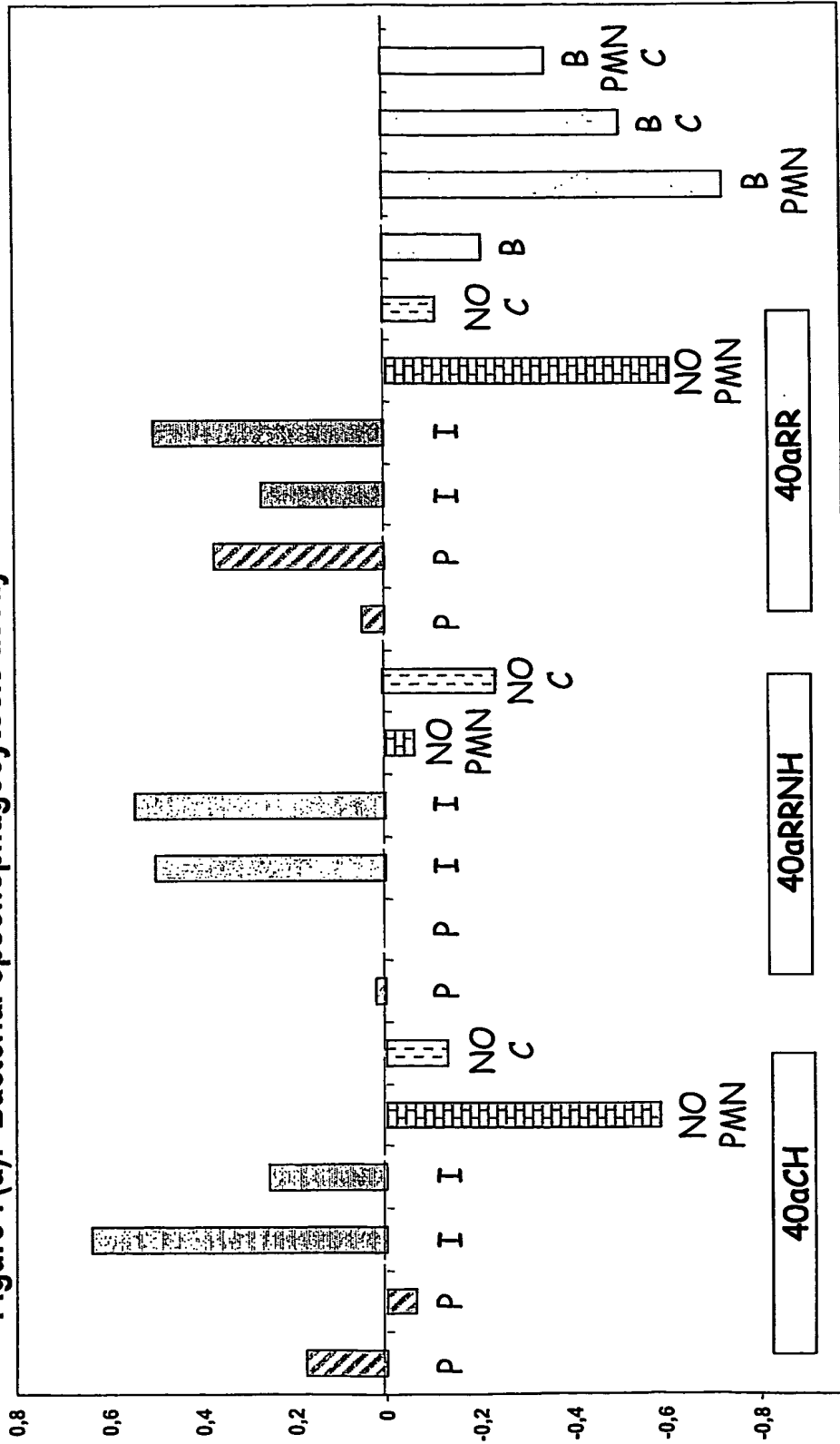
Figure 7(a): Bacterial opsonophagocytosis assay of GAS 40 constructs

B – bacteria
P – preimmune serum
I – immune serum
No PMN – without polymorphonucleates
No C – without complement

Histogram bars represent the difference between logarithm at T 0 (initial time) and T60 (bacterial CFU counted after 60 minutes of incubation)
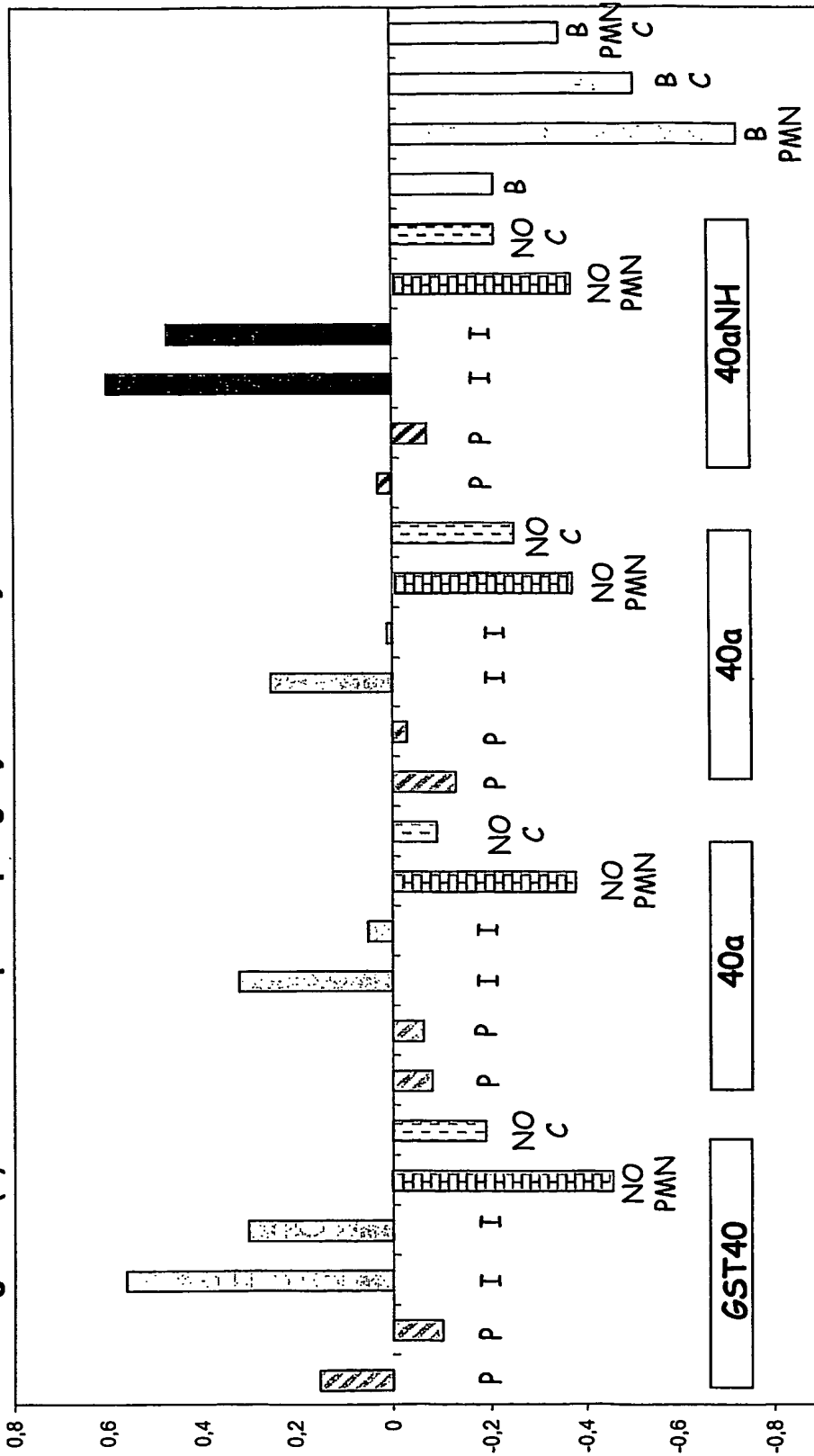
Figure 7(b): Bacterial opsonophagocytosis assay of GAS 40 constructs

B – bacteria
P – preimmune serum
I – immune serum
No PMN – without polymorphonucleates
No C – without complement

Histogram bars represent the difference between logarithm at T 0 (initial time) and T60 (bacterial CFU counted after 60 minutes of incubation)

Figure 8: Immunization in Murine Mouse Model

GAS antigen	Survival/Tested mice			Protection	pValue	Protein Purity
	alive	dead	tested	%	Chi-square	%
gst 40	67	63	130	51	0.000012	
253	14	36	50	28	0.006	15
253-urea	2	8	10	20		25
253-gst	2	8	10	20		30
39	9	31	40	22.5	0.09	20
39a	13	37	50	26	0.016	10
39a	10	30	40	25	0.039	
39a	12	28	40	30	0.0046	
urea 366	21	78	99	21.2	0.046	65
117	19	51	70	27	0.0036	15
117-urea	1	9	10	10		80
117-urea-2M	7	23	30	23.3	0.1	80
117-urea-2M (prep 117)	8	32	40	20	0.2	
urea 504	9	31	40	22.5	0.09	50
504	14	26	40	35	0.0003	40
504	7	33	40	17.5	0.4	80
urea 389	7	23	30	23	0.1	30
533	14	56	70	20	0.12	50
new 533	4	16	20	20	0.34	30
gst 57	12	48	60	20	0.14	60
57a	0	20	20	0		50
294	17	73	90	18.8	0.14	80
130	15	65	80	18.7	0.17	40
130	7	23	30	23.3	0.1	40
84	8	32	40	20	0.2	70
urea 159	7	33	40	17.5	0.4	5
159a	2	8	10	20		65
527	10	40	50	20	0.17	50
527	3	17	20	15		80
217	7	33	40	17.5	0.4	50
511	13	67	80	16.2	0.41	80
277	8	42	50	16	0.52	5
277a	2	28	30	6.6		50
gst 202	3	17	20	10	0.75	5
202a	5	25	30	16.6	0.53	5
45	5	25	30	16.6	0.53	80
urea 309	5	25	30	20	0.53	8
290	6	34	40	15	0.67	50
529	6	34	40	15	0.67	5
gst 58	10	60	70	14.2	0.71	30
384	7	43	50	14	0.78	80
384RR	1	19	20	5		80
urea 509	7	53	60	11.6	0.84	50
509-NH	2	8	10			75
509-CH	0	10	10			75
193	7	53	60	11.6	0.84	65
urea 372	4	25	29	13.7	0.85	20
gst 42	4	26	30	13.3	0.9	50
95	5	35	40	12.5	1	55
urea 236	5	35	40	12.5	1	80
new 236	2	8	10	20		70
137	5	35	40	12.5	1	75
His-Stop	29	201	230	12:06		